

Retesting the Hypothesis of a Clinical Randomized Controlled Trial in a Simulation Environment to Validate Anesthesia Simulation in Error Research (the VASER Study)

Alan F. Merry, M.B., Ch.B., Jacqueline A. Hannam, Ph.D., Craig S. Webster, Ph.D., Kylie-Ellen Edwards, M.B., Ch.B., Jane Torrie, M.B., Ch.B., Chris Frampton, Ph.D., Daniel W. Wheeler, Ph.D., Arun K. Gupta, Ph.D., Ravi P. Mahajan, D.M., Rachel Evley, Ph.D., Jennifer M. Weller, M.D.

ABSTRACT

Background: Simulation has been used to investigate clinical questions in anesthesia, surgery, and related disciplines, but there are few data demonstrating that results apply to clinical settings. We asked “would results of a simulation-based study justify the same principal conclusions as those of a larger clinical study?”

Methods: We compared results from a randomized controlled trial in a simulated environment involving 80 cases at three centers with those from a randomized controlled trial in a clinical environment involving 1,075 cases. In both studies, we compared conventional methods of anesthetic management with the use of a multimodal system (SAFERsleep®; Safer Sleep LLC, Nashville, Tennessee) designed to reduce drug administration errors. Forty anesthesiologists each managed two simulated scenarios randomized to conventional methods or the new system. We compared the rate of error in drug administration or recording for the new system *versus* conventional methods in this simulated randomized controlled trial with that in the clinical randomized controlled trial (primary endpoint). Six experts were asked to indicate a clinically relevant effect size.

Results: In this simulated randomized controlled trial, mean (95% CI) rates of error per 100 administrations for the new system *versus* conventional groups were 6.0 (3.8 to 8.3) versus 11.6 (9.3 to 13.8; $P = 0.001$) compared with 9.1 (6.9 to 11.4) versus 11.6 (9.3 to 13.9) in the clinical randomized controlled trial ($P = 0.045$). A 10 to 30% change was considered clinically relevant. The mean (95% CI) difference in effect size was 27.0% (-7.6 to 61.6%).

Conclusions: The results of our simulated randomized controlled trial justified the same primary conclusion as those of our larger clinical randomized controlled trial, but not a finding of equivalence in effect size. (**ANESTHESIOLOGY 2017; 126:472-81**)

SIMULATION has been used to investigate various clinical questions in anesthesia, surgery, and related disciplines. Some of this research has supported important changes in practice (*e.g.*, in relation to the use of checklists to manage crises).¹⁻⁴ Simulation offers several perceived advantages over clinical settings. Researchers can present the same standardized scenarios to all participants. This removes the confounding effects of differences between patients and procedures and allows paired statistical tests to be used. Prespecified clinical challenges can be included at will, effectively enriching “case mix” and obviating the need to study many routine cases to capture data related to infrequent events. If relevant, it is also possible to make errors more likely by increasing the complexity of the cases or by introducing latent factors thought to predispose to errors. Thus, smaller

What We Already Know about This Topic

- Simulation has been used to investigate clinical questions and for training in anesthesia
- There is, however, little evidence that results apply to clinical settings
- A previous randomized clinical trial evaluated a method to reduce error in drug administration during anesthesia

What This Article Tells Us That Is New

- This investigation repeated the clinical trial, but in a smaller simulated trial, to determine if the same principal conclusion would be reached as in the actual clinical trial
- The small simulated trial reached the same conclusion as the larger clinical trial, but the effect size was different
- Caution is needed when extrapolating findings from research in simulated settings to clinical practice

Submitted for publication June 10, 2016. Accepted for publication December 12, 2016. From the Department of Anaesthesiology, Faculty of Medical and Health Sciences (A.F.M., J.A.H., J.T.), Centre for Medical and Health Sciences Education, Faculty of Medical and Health Sciences (C.S.W., J.M.W.), University of Auckland, Auckland, New Zealand; Anaesthetic Department, St Mary's Hospital, London, United Kingdom (K.-E.E.); Department of Medicine, Christchurch School of Medicine and Health Sciences, University of Otago, New Zealand (C.F.); University Division of Anaesthesia, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom (D.W.W.); Department of Anaesthesia, Cambridge University Hospitals, Cambridge, United Kingdom (A.K.G.); and Anaesthesia and Critical Care, Division of Clinical Neuroscience, University of Nottingham, Nottingham, United Kingdom (R.P.M., R.E.).

Copyright © 2017, the American Society of Anesthesiologists, Inc. Wolters Kluwer Health, Inc. All Rights Reserved. Anesthesiology 2017; 126:472-81

sample sizes should be needed than in clinical settings, with potential savings in time and cost. Risks to patients are eliminated, and there should be no legal or regulatory risks to participants arising from the observation and recording of aspects of their clinical practice.⁵ The realism of simulation varies, and it seems probable that greater realism would enhance the transfer to clinical settings of results (in research) and points learned (in training). For these reasons, our team previously developed a highly realistic simulation model for research into human error in anesthesia,^{3,6} including error in drug administration.^{7,8} Although there is growing evidence to indicate that skills and behaviors learned in simulated settings can be expected to manifest in clinical practice,^{4,9,10} empirical data evaluating the external generalizability of conclusions drawn from simulation-based research to the clinical environment are scarce.^{6,11,12}

Therefore, we aimed to address the general question of whether the conclusions of research conducted using highly realistic simulation are likely to apply in clinical settings. To do this, we chose to retest a hypothesis evaluated in a previous, large, open-label, clinical randomized controlled trial (cRCT). In the cRCT, our group evaluated a multimodal initiative designed to reduce error in drug administration during anesthesia (“the new system”).¹³ The principal conclusion was that the new system was associated with a reduction in errors in the recording or administration of drugs. We sought to add to the evidence on the validity of using highly realistic simulation in evaluating interventions to improve safety in anesthesia by answering the specific question: Would the results of a smaller simulation RCT (sRCT) justify the same principal conclusion as those of this cRCT? We also undertook a *post hoc* analysis of the equivalence of the effect size of the intervention on the primary outcome variable.

Materials and Methods

We obtained approval for the study from local ethics committees (NTY/07/10/112 [Health and Disability Ethics Committee, Thorndon, New Zealand], 09/H0402/81 [Leicestershire, Northamptonshire and Rutland Research Ethics Committee 2, Northamptonshire, United Kingdom], and 09/H0308/126 [Cambridgeshire 2 Research Ethics Committee, Cambridge, United Kingdom]) and registered the study with the Australian and New Zealand Clinical Trials Registry (ACTRN12609000530224). Enrollment of participants at the first site began in March 2009, and data collection began 2 months later in May 2009. Participants gave written informed consent, and participants and researchers signed confidentiality agreements.

We defined our *binary* primary outcome as the acceptance or rejection of the hypothesis that there would be no significant difference between the new system and conventional methods in the rate of errors in the recording or administration of drugs, determined by (1) observation; (2)

reconciliation of anesthesia records with used ampules; and (3) information obtained at debriefing. This hypothesis was the same as that of the cRCT, so rejecting this hypothesis would justify the same principal conclusion as that reached in the cRCT and therefore add support to the validity of simulation in this context.

We also evaluated four other outcomes from the cRCT. Thus, the four secondary outcomes of the present sRCT were as follows: (1) actual drug administration error (assessed as for the primary outcome); (2) vigilance, assessed by the rate of lapses in responding to a vigilance latency task; (3) workload intensity measured by observation; and (4) time taken on record keeping, assessed by observation.

As in the cRCT, we also measured compliance with procedural principles of the new system and recorded the time allocated to certain tasks (to provide a comparison between simulation and clinical practice).

Study Design and Participants

This was a prospective RCT using simulation, replicating our previous cRCT as far as practicable while taking advantage of the potential benefits of simulation. Simulated cases were randomized to management of anesthesia using either the new system or conventional methods. In the cRCT, we observed a variety of cases from two adult operating suites at a single hospital. In the sRCT, we used two highly scripted study scenarios adapted from real clinical cases to promote a high level of replicability between cases (*i.e.*, the participants all managed the same two cases).

We collected data from three university-affiliated simulation centers: the University of Auckland’s Simulation Centre for Patient Safety (Auckland, New Zealand [NZ]); the Addenbrooke’s Postgraduate Education Centre (Cambridge, United Kingdom [U.K.]); and the Trent Simulation and Clinical Skills Centre, Queen’s Medical Centre (Nottingham, United Kingdom). We recruited 40 participants: 20 in Auckland, and 10 at each U.K. site. All senior trainee (“residents”) and specialist (“attending”) anesthesiologists working in the general anesthetic departments of tertiary teaching hospitals associated with the study centers were eligible to participate and were sent flyers advertising the project and inviting them to take part. The study was also presented to meetings of these departments. Those individuals who indicated an interest in participating and were available to participate on planned study days were provided more comprehensive written information sheets and asked to sign a consent form. Participants in Auckland were familiar with the new system, which had been in use in their hospital for more than 10 yr. Those in the United Kingdom had not used the new system before.

Randomization and Masking

Each anesthesiologist participated in two consecutive, simulated cases. The study statistician (C.F.) used computer-generated balanced random numbers (Microsoft Excel,

Redmond, Washington), to allocate cases to the new system or conventional methods and one of the two scenarios, with stratification for scenario order and morning *versus* afternoon. Each participant thus had one scenario with the new system and one with the conventional system. Blinding was not possible.

Intervention

The new system used in this study (SAFERsleep®; Safer Sleep LLC, Nashville, Tennessee) has been described previously.^{13,14} It includes a customized drug tray for organization of in-use syringes and vials; barcoded, color-coded drug labels; a touch-screen computer with barcode reader to provide auditory and visual cross-checking before drug administration; and real-time automated compilation of an anesthetic record. We provided prefilled syringes, labeled and color coded, for 12 commonly used anesthetic drugs. All color coding for the new system was according to an international color-code standard.¹⁵ Drug trolley drawers were also color coded and organized, so that drugs were grouped according to class and order of use during the conduct of routine anesthesia. One site (Cambridge, United Kingdom) had drug cupboards, which were reorganized for the new system cases to be consistent with the drug trolleys at the other sites. Participants viewed an instructional video on the new system and were asked to comply with six key procedural rules for its correct use (table 1).

For cases with conventional methods, participants made handwritten records on preprinted forms taken from their institutions (in Auckland, these were forms that predated the introduction of the new system). We provided the drug trays and labels (color coded to the same international standard) normally used in the participants' hospitals and their usual drug trolleys or cupboards. Participants were responsible for drawing up the drugs they wished to use into syringes from ampules and vials, and for labeling these syringes.

Data Collection

We collected data using the same methods as in the cRCT, and these are described more fully elsewhere.¹³ Briefly, we took an inventory of the anesthetic drugs provided for each case. We asked participants to retain all vials (empty or full) and syringes and not to discard unused contents of

Table 1. Procedural Rules Associated with Appropriate Use of the New System

Syringes and ampules should be organized in a tidy fashion. Ampules should be appropriately laid out at the start of the case.
All empty ampules and syringes should be retained for reconciliation.
All syringes should be labeled.
The computerized voice should be audible throughout each anesthetic.
The medication label should be scanned before each drug administration.

syringes. We provided empty sharps bins at the beginning of each case and inspected them later to identify any drugs discarded inadvertently during the case. From these sources, we worked out the drugs used and the total doses administered. As in the cRCT, we compared this information with that on the final anesthetic record and our observations to determine any discrepancies. We defined error in drug administration as (1) giving a drug other than the one intended (substitution) and (2) failing to give an intended drug (omission). We defined error in drug recording as (1) failing to record an administered drug; (2) failing to record the dose of a drug; or (3) a discrepancy between the recorded total dose and the observed total dose. Only discrepancies that were deemed clinically relevant by the criteria established in the cRCT were included.¹³ Where an inconsistency was found between the identities of the reconciled ampule, the labeled syringe, and the drug recorded by the anesthesiologist, the error was classified as "incorrect label" and categorized as an error in recording.

As in the cRCT,¹³ we assessed vigilance¹⁶ by asking participants to respond by touching a light illuminated on the screen of a personal digital assistant (HP iPAQ, Houston, Texas) positioned next to the screen of the anesthetic monitor. The light illuminated (and remained on until touched) at random intervals between 9 and 14 min, and time to response was recorded. We classed failure to respond within 300 s as a lapse of vigilance.

Observers assessed workload at random intervals of 7 to 15 min using the Borg workload scale.¹⁷ We recorded the time spent on prespecified intraoperative tasks during each case using custom task analysis software¹³ based on the study of Weinger *et al.*^{18,19} but modified to place emphasis on tasks related to drug administration and record making.

Observers assessed overall compliance with the six procedural rules of the new system (table 1) for each case. Possible scores for each rule were as follows: "yes" (2), "sometimes" (1), and "no" (0).

The Simulations

We aimed to make the simulation environment as realistic as possible and as consistent as possible between study sites, taking into account differences in local practices. We used a METI full-body human patient simulator (Medical Education Technology Inc., USA) in a simulated operating room at all sites. The roles of surgeons, scrub nurses, and circulating nurses were played by faculty familiar with the activities of these professions, but we provided a qualified anesthesia assistant in each case consistent with routine clinical practice (a technician in Auckland and an operating department assistant in the U.K.). We provided contemporary anesthesia equipment, including volatile anesthetics. We used real drugs in the case of ampules and vials, but substituted saline in realistically labeled ampules in the case of opioids. The prefilled syringes contained saline, but we used real propofol. The two case scenarios were highly scripted and adapted from

clinical cases. The cases included induction of anesthesia and reversal of anesthesia and muscle relaxation after the end of the simulated surgeries. We asked participants to behave as they would in normal clinical practice and provided enough time for them to complete the anesthetic record to their satisfaction before concluding each case.

In previous simulation-based research, we have used various means of increasing the likelihood of observing an error,⁷ and we employed two of these here. First, we made the scenarios clinically challenging but did not include any major life-threatening events. Second, we placed two ampules or prefilled syringes (depending on the arm of the study) of prespecified drugs into a compartment of the drug drawer (or cupboard) that contained another drug of a similar appearance. One of these incorrectly placed drugs was stored with another from the same drug class, the second with a drug of a different class.

Data Published Elsewhere

Data on other aspects of practice were collected during the NZ simulated cases and have been published previously.^{5,6,20,21}

Statistical Analysis

Data were analyzed using SPSS version 22.0 (SPSS, USA). A two-sided $P < 0.05$ was predefined as indicating statistical significance. Normality plots of data were used to confirm the appropriateness of using parametric models where relevant. We analyzed the rate of drug administration or recording errors per 100 administrations and the workload assessments using a general linear mixed model with study arm, participant and study site included as factors. The proportions of lapses in vigilance were analyzed with a McNemar chi-square test. We calculated total compliance scores for each case from scores for the individual principles of the new system and hence mean compliance for each participant. We did not correct P values for multiple testing. In a previous simulation-based study at the Auckland center, we identified a rate for all errors of 9.7 (SD 3.4) errors per case.⁷ On the basis of these data, we calculated that 21 participants would be required to demonstrate a 30% reduction in the error rate with a safety intervention, with 80% power and alpha of 0.05 (but see Discussion). We decided to recruit 20 participants at the NZ site and a further 20 at the two U.K. sites in view of the multiple site design to allow for any consequent reduction in the effect size or increase in variance. Our plan of analysis was revised in line with a request during peer review. We sent the published paper describing the cRCT to six independent experts (four local and two international) and asked them what effect size they would consider clinically important in relation to an initiative (such as ours) to reduce error. Our analysis then had two foci. First, we undertook a superiority analysis (designed *a priori*) comparing the two arms of the sRCT for which our study was adequately powered. As planned (and registered), we compared this result with that of the cRCT. We then undertook a *post hoc* equivalence analysis, applied after examination of the data, that compared the effect size on the primary outcome variable of the simulation trial with that of the clinical trial.

Results

Participants and Scenarios

We collected data during 40 study days, between May 2009 and August 2010. Forty anesthesiologists participated in the study, 15 (37.5%) senior trainees (with a minimum of 3 yr of experience of clinical anesthesia) and 25 (62.5%) specialist anesthesiologists. All had previously participated in simulation-based training or research at least once. Data were analyzed from 80 complete cases (40 per arm). The mean (95% CI) duration of the cases was 64 (61 to 67) min with the new system and 65 (62 to 67) minutes with conventional methods. The mean number of drugs administered was 11 per case with the new system *versus* 10 with conventional methods, and the mean number of bolus administrations was 17 with the new system *versus* 13 with conventional methods.

Primary Outcome

There were fewer errors in drug administration or recording with the new system than with conventional methods. Forty-one such errors were identified from 780 drug administrations with the new system (5.3%) *versus* 58 from 540 drug administrations with conventional methods (10.7%). The mean (95% CI) rate of these errors per 100 administrations calculated from the general linear model was 6.0 (3.8 to 8.3) with the new system and 11.6 (9.3 to 13.8) with conventional methods ($P = 0.001$; table 2). The mean (95% CI) of the difference was -5.5 (-8.7 to -2.3 ; table 2; with results from the cRCT displayed similarly for comparison). Site was not a significant factor in this analysis.

Secondary Outcomes

1. Errors in administering intravenous drugs: The mean (95% CI) rate of errors in drug administration (per 100 administrations calculated from the general linear model) was 0.81 (-0.0 to 1.6) with the new system *versus* 1.1 (0.3 to 1.9) with conventional methods ($P = 0.64$). The mean (95% CI) of the difference was -0.3 (-1.4 to 0.9; tables 2 and 3).
2. Vigilance: Data were not available on vigilance for one case because of equipment malfunction (so $n = 79$). Lapses in response occurred in 25% of cases with the new system and 26% with conventional methods ($P = 1.00$; McNemar chi-square).
3. Workload intensity: Mean (95% CI) Borg workload scores were 12.6 (12.3 to 12.9) with the new system and 12.3 (11.8 to 12.5) with conventional methods ($P = 0.071$).
4. Time taken on record keeping: Less time was spent on record keeping with the new system (8 m 43 s) than with the conventional methods (11 m 14 s, $P = 0.018$; table 4).

Comparison with the Previous Clinical Trial

The aforementioned results are shown in comparison with those from the cRCT in table 5, with conclusions for each trial in each case. We drew the same principal conclusion

Table 2. Rates of Error (95% CI) per 100 Administrations of Intravenous Drugs for the New System and Conventional Methods for the sRCT and the cRCT

Error Rates	Simulation-based RCT	Former Clinical RCT
Mean (SEM)	New system vs. conventional	New system vs. conventional
Errors in administration or recording	6.0 (3.8 to 8.3) vs. 11.6 (9.3 to 13.8) $P = 0.001$	9.1 (6.9 to 11.4) vs. 11.6 (9.3 to 13.9) $P = 0.045$
Mean diff. (95% CI)	<i>-5.5 (-8.8 to -2.3)</i>	<i>-2.4 (-4.8 to -0.0)</i>
RR	<i>0.5 (0.3 to 0.7)</i>	<i>0.9 (0.8 to 1.0)</i>
Drug administration		
Substitution	0.2 vs. 0.6	0.0 vs. 0.2
Omission	0.6 vs. 0.5	0.1 vs. 0.1
Total errors in drug administration	0.8 (-0.0 to 1.6) vs. 1.1 (0.3 to 1.9) $P = 0.6$	0.2 (0-0.3) vs. 0.3 (0.2 to 0.5) $P = 0.1$
Mean diff. (95% CI)	<i>-0.3 (-1.4 to 0.9)</i>	<i>-0.2 (-0.4 to 0.0)</i>
RR	<i>0.8 (0.2 to 3.4)</i>	<i>0.6 (0.3 to 1.4)</i>
Drug recording		
Labeling error	0.0 vs. 0.1	0.0 vs. 0.0
Drug and dose not recorded	0.6 vs. 4.3	2.3 vs. 3.5
Dose not recorded	0.0 vs. 0.6	0.1 vs. 0.7
Discrepancy in recorded dose	4.6 vs. 5.6	6.9 vs. 7.2
Total errors in drug recording	5.2 (3.1 to 7.4) vs. 10.5 (8.4 to 12.6)	9.7 (7.8 to 11.5) vs. 12.1 (10.1 to 14.1)
Mean diff. (95% CI)	<i>-5.3 (-8.3 to -2.2)</i>	<i>-2.4 (-4.7 to -0.1)</i>

There were 780 intravenous drug administrations with the new system, and 540 with conventional methods. Rates were calculated using a general linear mixed model (see Methods) so may differ from absolute error rates. The absolute difference in relative risk between the simulation randomized controlled trial and clinical randomized controlled trial for errors in administration or recording was 0.4. The absolute risk for errors in administrations alone was 0.2. Mean differences between arms (new system-conventional) with 95% CI are given in italics.

cRCT = clinical randomized controlled trial; Discrepancy in dose = clinically relevant discrepancy between the recorded total dose and that calculated through observation (using previously published criteria)¹³; omission = failing to give the intended drug; RCT = randomized controlled trial; RR = relative risk for an error in administration or recording with the new system versus conventional methods, calculated from the raw data (as opposed to adjusted rates derived from the general linear model); sRCT = simulation randomized controlled trial; substitution = incorrect drug given.

Table 3. Errors in the Administration of Intravenous Drugs by Intervention with the Number of Times Each Event Occurred in Parentheses

Error Type	New System	Conventional Methods
Substitution		
Interclass	Metaraminol for benzylpenicillin (1)	
Intraclass		Cefoxitin for cefazolin (1); pancuronium for suxamethonium (2)*
Omission	Amoxicillin plus clavulanic acid (1); glycopyrrolate (1); insulin (1); metronidazole (1), neostigmine (1)	Atracurium (1); metronidazole (1)
Total	6	5

There were 780 administrations with the new system and 540 with conventional methods.

*Substitution errors resulting from an ampule deliberately placed incorrectly in trolley or cupboard with similar looking ampules by research staff (see Methods). Note: there was one "labeling error" in which cefazolin was labeled as cefoxitin, in the conventional arm. Although we consider this to be a drug administration error, we included this as a recording error (as in our previous study) as it is unclear whether the anesthetist's intention was to give the drug that was recorded, labeled on the syringe or actually in the syringe.

in the sRCT as we did in the cRCT, namely, the use of the new system was associated with a reduction in the rate of error in the administration or recording of drugs.

Conclusions drawn from three of the four evaluated secondary outcomes (errors in drug administration, workload,

and mean time spent recording with the new system) were also the same for the sRCT as for the cRCT. In the cRCT, there was technically no difference between groups for vigilance when judged on P value alone, but we were reluctant to rule out the possibility of a type II error with a P value of 0.052. In the sRCT, the result was clearly nonsignificant. Thus, our conclusions for vigilance were not the same between studies.

Post hoc Equivalence Analysis. Five of the six consulted experts responded. Their opinions of a clinically relevant effect size (*i.e.*, change in error rate) ranged from 10 to 30%. As a percentage of the primary outcome variable in the conventional arm of the cRCT (11.6 errors in administration or recording per 100 administrations), the mean (95% CI) difference in effect size between the simulation and clinical studies was 27.0% (-7.6 to 61.6%), with the larger effect size seen in the sRCT. The highest estimate of a clinically important effect size (30%) falls within this CI, so equivalence could not be demonstrated (fig. 1).

Other Data

Participants were fully compliant with all principles of the new system in only 5% of all cases (compared with 18% in the cRCT). Mean scores for compliance were 75.5% overall (SD 15.2%, range, 40 to 100%). They were 82.3% for the NZ site (SD 9.5%, range, 60 to 100%) and 68.8% for the U.K. sites (SD 17.0%, range, 40 to 100%). Information on duration of anesthesia in the two studies, and on the times spent on various tasks, is summarized in table 4.

Table 4. Duration of Anesthesia and Mean (SD) Times Spent on Various Tasks in the Present sRCT and in the Previous cRCT

	New System vs. Conventional	
	sRCT	cRCT
Duration of anesthesia, h		
< 1	100 vs. 100%	48 vs. 46%
1–2		34 vs. 36%
2–3		11 vs. 10%
> 3		7 vs. 8%
Task		
Recording	8 m 43 s vs. 11 m 14 s (4 m 32 s)* <i>P</i> = 0.018 <i>Less time spent recording with the new system</i>	10 m 23 s (9 m 51 s) vs. 20 m 5 s (10 m 6 s) <i>P</i> < 0.001 <i>Less time spent recording with the new system</i>
Drug preparation*	7 m 25 s vs. 12 m 43 s (4 m 26 s)	15 m 30 s (6 m 58 s) vs. 14 m 23 s (6 m 58 s)
Drug administration†	4 m 45 s vs. 4 m 42 s (2 m 51 s)	9 m 2 s (4 m 20 s) vs. 8 m 2 s (4 m 20 s)
Airway procedures	5 m 55 s vs. 4 m 57 s (2 m 44 s)	9 m 45 s (9 m 8 s) vs. 9 m 43 s (9 m 22 s)
Adjusting equipment	1 m 48 s vs. 2 m 29 s (1 m 48 s)	10 m 12 s (7 m 56 s) vs. 9 m 47 s (7 m 56 s)
Observing monitors	19 m 10 s vs. 21 m 18 s (12 m 39 s)	20 m 22 s (15 m 9 s) vs. 23 m 1 s (15 m 23 s)
Observing patient	6 m 24 s vs. 5 m 57 s (3 m 41 s)	28 m 26 s (20 m 11 s) vs. 21 m 22 s (20 m 26 s)
Tidying work space	37 s vs. 43 s (25 s)	3 m 28 s (2 m 24 s) vs. 3 m 17 s (2 m 24 s)
Task-related conversation‡	26 m 41 s vs. 28 m 15 s (10 m 1 s)	33 m 42 s (18 m 59 s) vs. 31 m 27 s (19 m 14 s)
Task-related reading	1 m 58 s vs. 2 m 23 s (1 m 35 s)	1 m 41 s (2 m 53 s) vs. 1 m 57 s (3 m 7 s)
Other tasks	1 m 18 s vs. 50 s (57 s)	6 m 10 s (7 m 27 s) vs. 7 m 1 s (7 m 27 s)

Statistical testing was only undertaken for recording because this was the only prespecified outcome supported by a hypothesis. Durations were calculated using a general linear model. Conclusions are shown in italics.

*Drug preparation included preparing intravenous drugs for administration by bolus or infusion. †Administration of intravenous drugs by bolus or infusion and adjusting vaporizers and gas flow. The following tasks were recorded in the cRCT but were not required in the sRCT: establishing vascular access and establishing/adjusting intravenous infusions; positioning the patient. ‡In the cRCT task-related conversation was combined with teaching-related conversation; the latter was not relevant in this simulation study because each case was carried out by a single anesthesiologist with no opportunity for teaching a resident.

cRCT = clinical randomized controlled trial; sRCT = simulation randomized controlled trial.

Table 5. Comparison of the Results and Conclusions of the Primary and Selected Secondary Outcomes between the Present sRCT and the Previous cRCT

Outcomes	New System vs. Conventional	
	sRCT	Former cRCT
Primary		
Mean difference (95% CI) in errors in the recording or administration of intravenous drugs per 100 drug administrations	–5.5 (–8.8 to –2.3) with new system (<i>P</i> = 0.001) <i>Null hypothesis rejected – fewer errors with the new system</i>	–2.4 (–4.8 to –0.0) with new system (<i>P</i> = 0.045) <i>Null hypothesis rejected – fewer errors with the new system</i>
Secondary		
Mean difference (95% CI) in errors in the administration of drugs per 100 drug administrations	–0.3 (–1.4 to 0.9) (<i>P</i> = 0.64) <i>No significant difference</i>	–0.18 (–0.4 to 0.0) (<i>P</i> = 0.12) <i>No significant difference</i>
Vigilance: percentage of cases in which a lapse occurred	25% vs. 26% (<i>P</i> = 1.00) <i>No difference</i>	12% vs. 9% (<i>P</i> = 0.052) <i>Technically no difference but discussion acknowledged that more lapses occurred with the new system</i>
Mean (95% CI) workload	12.6 (12.3 to 12.9) vs. 12.7 (11.8 to 12.5) (<i>P</i> = 0.071). <i>No difference between groups</i>	13.2 (12.8 to 13.6) in each group (<i>P</i> = 0.967) <i>No difference between groups</i>
Mean (SD) time spent recording	8 m 43 s vs. 11 m 14 s (4 m 32 s) (<i>P</i> = 0.018) <i>Less time recording with the new system</i>	10 m 23 s (9 m 51 s) vs. (10 m 6 s) (42 s) (<i>P</i> < 0.001) <i>Less time recording with the new system</i>

The difference between groups in the primary outcome variable was greater in the sRCT but the confidence limits were tighter in the cRCT, reflecting the relative sizes of the studies (40 vs. 500 cases per group, respectively). Conclusions are shown in italics.

cRCT = clinical randomized controlled trial; sRCT = simulation randomized controlled trial.

Discussion

The results of our sRCT in a highly realistic simulated environment justified the same principal conclusion as those of our larger cRCT and add some support to the generalizability to clinical settings of findings from simulation-based research for

evaluating safety initiatives in anesthesia. The data also justify the same conclusions for three out of four secondary outcomes evaluated in both studies (table 5). In both studies, the results add to evidence supporting the safety principles embedded in the new system. The mean effect size was greater in the sRCT

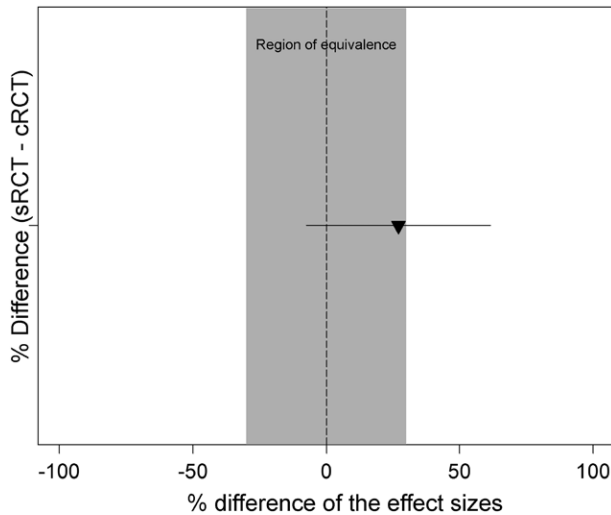


Fig. 1. The difference (%) in the effect sizes of the two studies; the region of equivalence is based on the assumption that a difference of 30% would be clinically relevant. cRCT = clinical randomized controlled trial; sRCT = simulation randomized controlled trial.

than in the cRCT. As can be seen from figure 1, the mean (95% CI) of the difference does not support a conclusion of equivalence for these results. It should be noted that the sample sizes of the two studies were based on their planned primary superiority analyses, and our evaluation of equivalence was undertaken *post hoc*. Given the size of our sRCT, our analysis of equivalence is inconclusive (see discussion under Strengths and Weaknesses of this Study). Our experts' view of clinical relevance ranged from 10 to 30%. We chose to use 30% as the clinically relevant difference for our equivalence analysis. It would have made no difference to our conclusions had we chosen 10%.

Validity (considered as a unified concept) should be supported with evidence accumulated from multiple sources.²² Our results add to evidence on the validity of simulation for certain types of research in anesthesia.^{6,23–25} Not all questions lend themselves to simulation, but simulation has been used in several studies on important aspects of safety in the operating room.^{1–3,26–31} Our group's recent study, which compared patterns of communication during routine anesthesia in simulated and clinical settings,⁶ focused on the validity of simulation in education, but its findings also have relevance to research. Our current study breaks new ground by focusing on the validity of simulation as a tool for research.

The difference between the sRCT and cRCT in respect of vigilance (the secondary outcome for which the conclusions drawn differ) is difficult to explain. One possible explanation includes greater reliance on monitors in both arms of the sRCT because clinical signs are of little or no use in simulation (noting that the latency task was adjacent to the monitors).

Strengths and Weaknesses of This Study

We have developed and refined our simulation model during several years of research and teaching centered around piloting and evaluating safety initiatives in anesthesia and teaching

and investigating human factors and team behaviors.^{3,7,8,14} Inevitably there are limits to the degree of realism that can be achieved with anesthetic simulators that are currently available commercially, but we put emphasis on achieving as high a level of realism as possible.^{6–8,24} The realism of our simulations is a strength, but our findings may not extend to simulation-based research in which the level of realism is lower. We were surprised that there were 780 drug administrations with the new system and 540 drug administrations with conventional methods. Perhaps automatic recording of physiologic data promotes earlier and more frequent pharmacologic interventions³²—but this is speculation, and we do not know what led to this difference. Although the denominator is an important factor in the determination of the rate of error in each group, fewer drug administrations would be associated with fewer opportunities for error, so we do not think this difference is likely to have affected our conclusions unduly. Another limitation lies in our sampling, which was dictated by convenience—primarily the availability of our participants. This weakness was offset by our multicenter design. We took care to minimize differences and align research practices between the sites as much as possible, with extensive observer training and site visits of key personnel during study days. Notably, Auckland participants were familiar with the new system while U.K. participants were not (this is reflected in the compliance scores, mean score of 82% for the NZ site *vs.* 69% for the U.K. sites), and this is a possible source of bias. However, the mean compliance score for the sRCT (76%) did not differ greatly from that in the cRCT (81%). Furthermore, we would expect any impact of novice users in the sRCT to reduce the effect size of the intervention, whereas it was actually greater in the sRCT than in the cRCT. Provision of assistance to the anesthesiologists was different between sites (this was by operating department assistants in the U.K. and anesthetic technicians in NZ). This may have altered the distribution of task allocation in the sRCT. An important limitation is the *post hoc* nature of the test for equivalence. An analysis of equivalence, planned *a priori*, would be a more robust statistical approach to a comparison of two research methodologies. However, to adequately power such an analysis, we would have needed approximately 1,000 simulations (500 per study arm), had we assumed that a 30% difference was clinically meaningful and used estimates of the rate of error and variance from the cRCT. A sRCT of this size would not have been affordable or practical and would negate a key point about simulation, namely, that the simulation-based design we adopted makes it possible to address superiority hypotheses with much smaller studies than in a clinical setting.

Strengths and Weaknesses of Simulation as a Research Tool in General

In addition to the advantages and disadvantages of simulation outlined in the introduction, we also note that all recruited participants took part in all cases. We have found this to be typical of simulation-based studies but in contrast with our cRCT,

in which even participants who had agreed to take part in the study declined to participate on the day in more than 150 cases. Debriefing, which would be difficult in large clinical studies, is facilitated in simulation and allows, among other things, uncertainties in the data to be clarified. An obvious disadvantage of simulation is the inability to evaluate patient outcomes.

One advantage of simulation is the potential to increase statistical power. Our sRCT required only 80 cases in comparison to 1,075 cases in our cRCT. However, there is a trade-off—larger numbers do provide greater surety that unknown sources of bias will be distributed evenly between groups through the process of randomization, and confidence limits will be narrower. Thus, we still consider a large, well-designed clinical RCT to be the accepted standard for answering questions that lend themselves to this design—but simulation may offer a viable alternative, or perhaps a starting point, for investigating certain research questions. In general, simulation lends itself to studies of aspects of human factors, teamwork, and communication, but not to studies that depend on the biologic functions of the subjects, such as investigations into pharmacology or physiology. It is also not possible to investigate the impact of an intervention on patient outcomes using simulation. The value of triangulation between evidence from different studies should be kept in mind. It may sometimes make economic and practical sense for simulation to be used in obtaining a preliminary answer to a research question, while at the same time informing the design of a subsequent confirmatory clinical trial, and the robustness of aligned findings from two different approaches may exceed that of either approach alone.

The data in table 4 reflect certain similarities and differences between simulated and clinical anesthetic settings. The simulated cases were all of similar duration, whereas the clinical cases varied considerably in duration (and presumably in complexity). This may imply that findings from the cRCT are more generally applicable: the experience of managing very short or very long anesthetics may be substantially different from that of managing a standardized half-hour case. This also explains some of the absolute differences between times spent on activities in the two studies.

The Safety Principles Underpinning the Anesthesia Information Management System

The new system incorporates a multifaceted approach to reducing errors in the administration of drugs and facilitates recording of the relevant information.¹⁴ Many of the principles of this system are aligned with those underpinning the “New Paradigm” proposed by the Anesthesia Patient Safety Foundation (Indianapolis, Indiana) for drug administration in anesthesia.³³ Our results add to the evidence supporting these principles, even though (as with the cRCT) compliance with the use of these principles left much to be desired. In both studies, the primary outcome included errors in recording and in administering drugs. This reflects the very large sample size that would be needed to evaluate errors in administration alone, but, as explained in the cRCT, “The anesthetic record is

an important clinical tool for decision making and a source of data for research, audit, legal purposes, and continuous quality improvement. Inaccuracies in the recording of administered drugs can lead to subsequent errors, such as the repeated administration of a dose of drug given but not noted.”

Conclusions

The results of our sRCT justified the same principal conclusion and also several secondary conclusions as those of our larger cRCT. This adds to evidence supporting the principles of the new system and also provides some support for the use of simulation in the investigation of applicable research questions in anesthesia, at least in respect of highly realistic simulation. On the other hand, we were not able to demonstrate equivalence, and the differences in effect size between the two studies may reflect the well-known differences between simulated and clinical environments. Thus, caution is still needed when extrapolating findings from research in simulated settings to clinical practice.

Acknowledgments

The authors thank the anesthesiologists and anesthetic technicians who participated in this study. We also acknowledge the following individuals for their contributions: Bryn Baxendale (M.B., Ch.B., F.R.C.A., Director, Trent Simulation and Clinical Skills Centre, Nottingham, United Kingdom) for supervision and contribution to the study days; Mark Kane (Senior Technician, Trent Simulation and Clinical Skills Centre, Nottingham, United Kingdom) for technical support and contribution to study days; David L. Evans (F.R.C.A.), Louise Murray, Tracey-Ann Reader, Charles D. Hallett (Simulation Centre, Postgraduate Medical Centre, Addenbrooke’s Hospital, Cambridge, United Kingdom), Claire E. Williams (B.M., Ch.B., F.R.C.A., Department of Anaesthesia, Addenbrooke’s Hospital, Cambridge, United Kingdom), and Kaylene Henderson (Project Manager, Centre for Medical and Health Sciences Education, University of Auckland, Auckland, New Zealand) for contribution to study days; Derryn Gargiulo (M.Pharm., Department of Anaesthesiology, University of Auckland, Auckland, New Zealand) for simulated prefilled syringe manufacture; Robert Henderson for oversight with particular reference to the previous clinical study (M.Sc., Human Factors Consultant, Human Factors Group, Simulation Training, Air New Zealand, Auckland, New Zealand); Dr. Lara Hopley (M.B., Ch.B., Department of Anaesthesia and Perioperative Medicine, Waitemata District Health Board, Auckland, New Zealand) for participation on study trial run days; Dr. Anisoara Jardim (Ph.D., National Institute for Health Research Central Commissioning Facility, London, United Kingdom) for data collection; David Merry (M.A., Department of Philosophy, Humboldt University, Berlin, Germany) for writing the task allocation and vigilance software; Dr. Matthew Weinger (M.D., Professor, Department of Anesthesiology, Vanderbilt University Medical Centre, Tennessee) for permission and assistance with task allocation and workload assessments; and the Patient Safety research group at the University of Auckland for their various contributions to study days; and the following experts for advice on a clinically relevant difference in the rate of error in the recording and administration of drugs: Drs. David Bates (M.D., Department of Medicine, Brigham and Women’s Hospital, Boston, Massachusetts), Jeremy Cooper (M.B., Ch.B., Department of Cardiac and ORL Anaesthesia, Auckland District Health Board, Auckland,

New Zealand), Michael Gilham (M.B.B.S., Cardiothoracic and Vascular Intensive Care and High Dependency Unit, Auckland District Health Board, Auckland, New Zealand), Simon Mitchell (Ph.D., Associate Professor, Department of Anaesthesiology, University of Auckland, Auckland, New Zealand), and Joyce Wahr (M.D., Professor, Department of Anesthesiology, University of Minnesota, Minneapolis, Minnesota).

Research Support

Supported by project grants from the Australian and New Zealand College of Anaesthetists (Melbourne, Victoria, Australia), the Auckland Medical Research Foundation (Auckland, New Zealand), and the National Institute for Academic Anaesthesia (London, United Kingdom).

Competing Interests

All authors have completed the International Committee of Medical Journal Editors uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare that A. F. Merry and Dr. Webster are shareholders in Safer Sleep (SAFERsleep®; Safer Sleep LLC, London, United Kingdom). A. F. Merry is a director of Safer Sleep, holds about 9% of its shares, and advises the company on the design of its products. Dr. Webster is a minor shareholder. The intellectual property of the new system is owned by Safer Sleep, although some patents are in the name of A. F. Merry (as inventor). A. F. Merry and Dr. Webster have been authors on several previous publications evaluating the new system used in this research. The other authors declare no competing interests.

Correspondence

Address correspondence to A. F. Merry: Department of Anaesthesiology, School of Medicine, University of Auckland, Private Bag 92019, Auckland, New Zealand. a.merry@auckland.ac.nz. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY'S articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

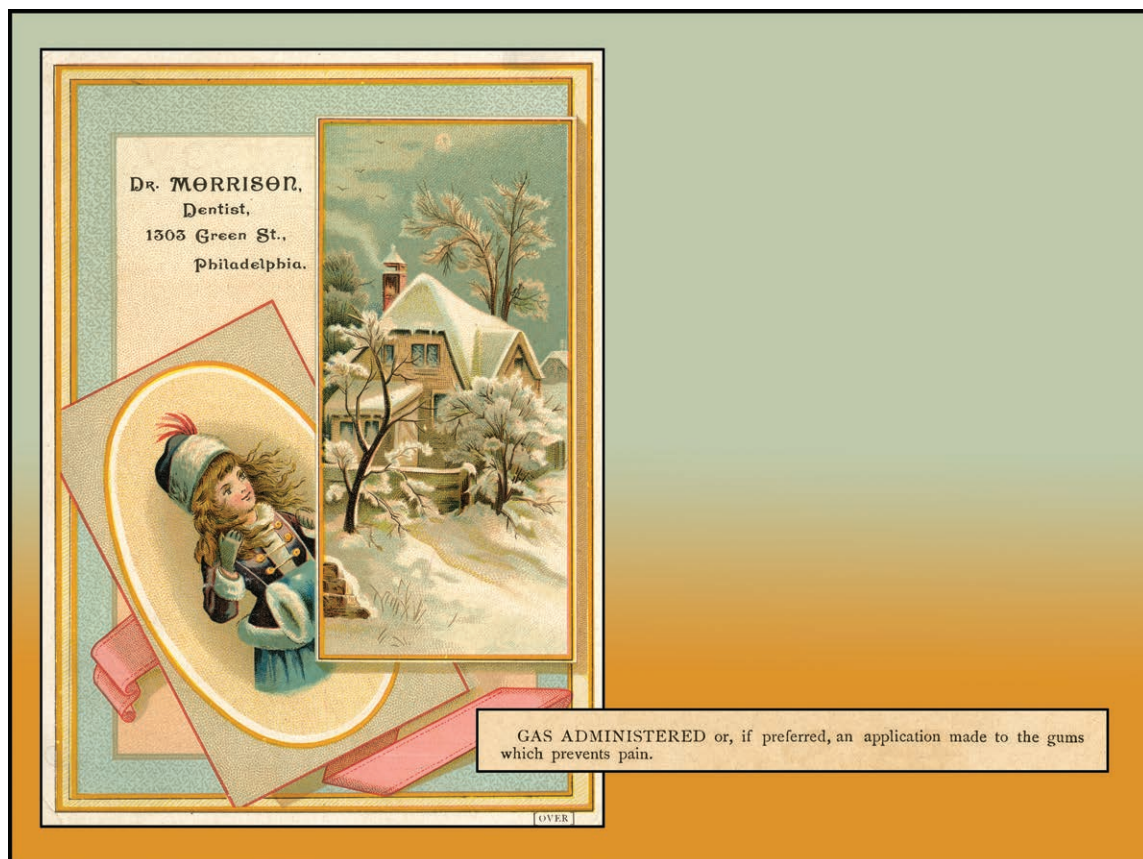
- Arriaga AF, Bader AM, Wong JM, Lipsitz SR, Berry WR, Ziewacz JE, Hepner DL, Boorman DJ, Pozner CN, Smink DS, Gawande AA: Simulation-based trial of surgical-crisis checklists. *N Engl J Med* 2013; 368:246–53
- Weller J, Merry A, Warman G, Robinson B: Anaesthetists' management of oxygen pipeline failure: Room for improvement. *Anaesthesia* 2007; 62:122–6
- Weller JM, Merry AF, Robinson BJ, Warman GR, Janssen A: The impact of trained assistance on error rates in anaesthesia: A simulation-based randomised controlled trial. *Anaesthesia* 2009; 64:126–30
- Weller J, Frengley R, Torrie J, Webster CS, Tomlinson S, Henderson K: Change in attitudes and performance of critical care teams after a multi-disciplinary simulation-based intervention. *Int J Med Educ* 2012;3:124–31.
- Webster CS, Andersson E, Edwards K, Merry AF, Torrie J, Weller JM: Deviation from accepted drug administration guidelines during anaesthesia in twenty highly realistic simulated cases. *Anaesth Intensive Care* 2015; 43:698–706
- Weller J, Henderson R, Webster CS, Shulruf B, Torrie J, Davies E, Henderson K, Frampton C, Merry AF: Building the evidence on simulation validity: Comparison of anaesthesiologists' communication patterns in real and simulated cases. *ANESTHESIOLOGY* 2014; 120:142–8
- Merry AF, Weller JM, Robinson BJ, Warman GR, Davies E, Shaw J, Cheeseman JF, Wilson LF: A simulation design for research evaluating safety innovations in anaesthesia. *Anaesthesia* 2008; 63:1349–57
- Merry AF, Webster CS, Weller J, Henderson S, Robinson B: Evaluation in an anaesthetic simulator of a prototype of a new drug administration system designed to reduce error. *Anaesthesia* 2002; 57:256–63
- McGaghie WC, Draycott TJ, Dunn WF, Lopez CM, Stefanidis D: Evaluating the impact of simulation on translational patient outcomes. *Simul Healthc* 2011; 6(Suppl):S42–7
- Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hamstra SJ: Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *JAMA* 2011; 306:978–88
- Ross AJ, Kodate N, Anderson JE, Thomas L, Jaye P: Review of simulation studies in anaesthesia journals, 2001–2010: Mapping and content analysis. *Br J Anaesth* 2012; 109:99–109
- Rosen KR: The history of medical simulation. *J Crit Care* 2008; 23:157–66
- Merry AF, Webster CS, Hannam J, Mitchell SJ, Henderson R, Reid P, Edwards KE, Jardim A, Pak N, Cooper J, Hopley L, Frampton C, Short TG: Multimodal system designed to reduce errors in recording and administration of drugs in anaesthesia: Prospective randomised clinical evaluation. *BMJ* 2011; 343:d5543
- Merry AF, Webster CS, Mathew DJ: A new, safety-oriented, integrated drug administration and automated anaesthesia record system. *Anesth Analg* 2001; 93:385–90
- The International Organization for Standardization. Anaesthetic and respiratory equipment—User-applied labels for syringes containing drugs used during anaesthesia—Colours, design and performance. ISO 26825:2008, Switzerland, 2008.
- Weinger MB, Reddy SB, Slagle JM: Multiple measures of anaesthesia workload during teaching and nonteaching cases. *Anesth Analg* 2004; 98:1419–25
- Borg G, editor: *Simple Rating Methods of Perceived Exertion*. Oxford, Pergamon Press, 1977
- Weinger MB, Herndon OW, Gaba DM: The effect of electronic record keeping and transesophageal echocardiography on task distribution, workload, and vigilance during cardiac anaesthesia. *ANESTHESIOLOGY* 1997; 87:144–55
- Weinger MB, Herndon OW, Zornow MH, Paulus MP, Gaba DM, Dallen LT: An objective methodology for task analysis and workload assessment in anaesthesia providers. *ANESTHESIOLOGY* 1994; 80:77–92
- Houliston BR, Parry DT, Merry AF: TADAA: Towards automated detection of anaesthetic activity. *Methods Inf Med* 2011; 50:464–71
- Gargiulo DA, Sheridan J, Webster CS, Swift S, Torrie J, Weller J, Henderson K, Hannam J, Merry AF: Anaesthetic drug administration as a potential contributor to healthcare-associated infections: A prospective simulation-based evaluation of aseptic techniques in the administration of anaesthetic drugs. *BMJ Qual Saf* 2012; 21:826–34
- Kane MT: Current concerns in validity theory. *J Educ Meas* 2001;38(4):319–42
- Cumin D, Weller JM, Henderson K, Merry AF: Standards for simulation in anaesthesia: Creating confidence in the tools. *Br J Anaesth* 2010; 105:45–51
- Merry AF, Weller JM, Robinson BJ, Warman GR, Davies E, Shaw J, Cheeseman JF: A simulation design for research evaluating safety innovations in anaesthesia. *Anaesthesia* 2008; 63:1349–57
- Rudolph JW, Simon R, Raemer DB: Which reality matters? Questions on the path to high engagement in healthcare simulation. *Simul Healthc* 2007; 2:161–3
- Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R: Assessment of clinical performance during simulated crises

using both technical and behavioral ratings. *ANESTHESIOLOGY* 1998; 89:8–18

27. Gawande AA, Arriaga AF: A simulation-based trial of surgical-crisis checklists. *N Engl J Med* 2013; 368:1460
28. Howard SK, Gaba DM, Smith BE, Weinger MB, Herndon C, Keshavacharya S, Rosekind MR: Simulation study of rested *versus* sleep-deprived anesthesiologists. *ANESTHESIOLOGY* 2003; 98:1345–55
29. Kennedy RR, Merry AF, Warman GR, Webster CS: The influence of various graphical and numeric trend display formats on the detection of simulated changes. *Anaesthesia* 2009; 64:1186–91
30. Mudumbai SC, Fanning R, Howard SK, Davies MF, Gaba DM: Use of medical simulation to explore equipment failures and human-machine interactions in anesthesia machine pipeline supply crossover. *Anesth Analg* 2010; 110:1292–6
31. Weller JM, Janssen AL, Merry AF, Robinson B: Interdisciplinary team interactions: A qualitative study of perceptions of team function in simulated anaesthesia crises. *Med Educ* 2008; 42:382–8
32. van Schalkwyk JM, Lowes D, Frampton C, Merry AF: Does manual anaesthetic record capture remove clinically important data? *Br J Anaesth* 2011; 107:546–52
33. Eichhorn J. APSF hosts medication safety conference: Consensus group defines challenges and opportunities for improved practice. *APSF Newsletter* 2010;25(1):1–7.

ANESTHESIOLOGY REFLECTIONS FROM THE WOOD LIBRARY-MUSEUM

Philadelphia's Dr. W. S. Morrison: Advertising Winter's Chill... and Vapocoolant Application?



In February 1881, “W. S. Morrison” of Iowa received his D.D.S. degree from the Philadelphia Dental College. As winters began in 1882 and then in 1883, Dr. William S. Morrison advertised in *The Times* of Philadelphia that there was “No charge for gas if teeth are to be inserted” and then that he filled “teeth without pain by use of medicines, exercise of gentleness and care.” Continuing his winter advertising emphasis, Morrison released a trade card (*left*) featuring snow-covered trees and homes and a young girl dressed in a fur-trimmed hat, coat, and muffler. There is some chance that, in these days before cocaine’s availability, his “application made to the gums” involved chilling teeth to numbness with a vapocoolant such as an ether spray. A vapocoolant might also explain his preference for wintry motifs on this trade card from the Wood Library-Museum’s Ben Z. Swanson Collection. (Copyright © the American Society of Anesthesiologists’ Wood Library-Museum of Anesthesiology.)

George S. Bause, M.D., M.P.H., *Honorary Curator and Laureate of the History of Anesthesia, Wood Library-Museum of Anesthesiology, Schaumburg, Illinois, and Clinical Associate Professor, Case Western Reserve University, Cleveland, Ohio.* UJYC@aol.com.